

OpenBioMaps data management service for biological sciences and biodiversity conservation pilot

Principal Investigators: [Miklos Ban](#)

Shepherds: [Miguel Caballer](#)

Entry in the community requirement database: [OpenBioMaps data management service for biological sciences and biodiversity conservation](#)

About the pilot

We would like to create a service with EOSC that allows multiple users to run tasks that are above the level of a PC through the same interface. In fact, we would like to develop a "service in service" - specifically for projects that collect nature conservation and biodiversity data.

The most common computing tasks come from the following areas:

- Spatial analyses (satellite image processing, large or fine scale spatial queries),
- ML analyses (satellite image analyses, drone image analyses, distribution models (Random Forest), population dynamics analyses, survival analyses, supervised and non-supervised image classification - photos of individual animals, habitats, survey methods),
- Conservation genetics analyses.

To serve these diverse tasks we need a fully configurable VM which let us deploy our service interface (API) which will be available in the OpenBioMaps Network and provide computation capacity access to the involved projects.

According to our recent experiences in our PC based local computational cluster, the number of processors is the most important in these ecological analyzes. A "typical" analysis is now running at an acceptable rate on 16 threads. The parallel computing requirements of image analysis can be much higher, and GPU usage can be interesting there. Some analyzes, for example, genetic analyzes or larger spatial analyzes require a lot of memory.

Description of supported work

General

- We would like to develop a background service that supports the interpretation of data from databases on conservation biology and biodiversity. A solution that facilitates and generalizes the most common high-computational analysis of data stored in such databases.
- The OBM system is used for data management by nature conservation institutes, biodiversity research and citizen science projects. OBM provides a number of services that make day-to-day work with data easier, but we do not yet provide tools for analyzing the data. Particularly for high computing tasks.

We would like to create an interface that allows user-level access to data computing from OBM-based databases and to run custom analyzes so that the system supports users in the preparation of new analyzes based on the analysis performed.

Use Cases

- <https://openbiomaps.org/projects/>

Team

Participant	Role	Name and Surname
UNIDEB	PI	Miklós Bán
UPV	Shepherd	Miguel Caballer
	Technical support	

Technical Plan

The full technical plan can be found here:

Work planned for Q1	<ul style="list-style-type: none">• Integration with EGI Cloud Compute.• Deploy and Configure OBM node in test environment manually.
---------------------	---

Work planned for Q2	<ul style="list-style-type: none"> • Create TOSCA Recipes and Ansible roles needed to deploy the application automatically using IM.
Work planned for Q3	<ul style="list-style-type: none"> • Deploy OBM node to production environment using the developed recipes. • Analyse EOSC data services to be used by the application: <ol style="list-style-type: none"> 1. EGI DataHub. 2. B2 services (Drop, Find, Handle, Share) 3. EGI Services (Training Infrastructure, Data Transfer) 4. EOSC Marketplace services (GeoDAB, D4Science spatial services, Alien and Invasive Species Virtual Research Environment, Biodiversity, EODC JupyterHub for global Copernicus data)
Work planned for Q4	<ul style="list-style-type: none"> • Performance test of all nodes.

EOSC services and providers

Providers

- EGI: SLA and OLA with IFCA: <https://documents.egi.eu/public/ShowDocument?docid=3565>

Services

- EGI Cloud Compute
- Infrastructure Manager

Services that will be explored during project lifetime::

- GeoDAB
- D4Science spatial services
- Alien and Invasive Species Virtual Research Environment
- Biodiversity
- EODC JupyterHub for global Copernicus data